SKILLSYNC: The AI-Powered Job Search Engine

ETL Pipeline & RAG Model Development

Team Members:
Sanath Kumar Bajjuri
Pushpa Ugma Vemula
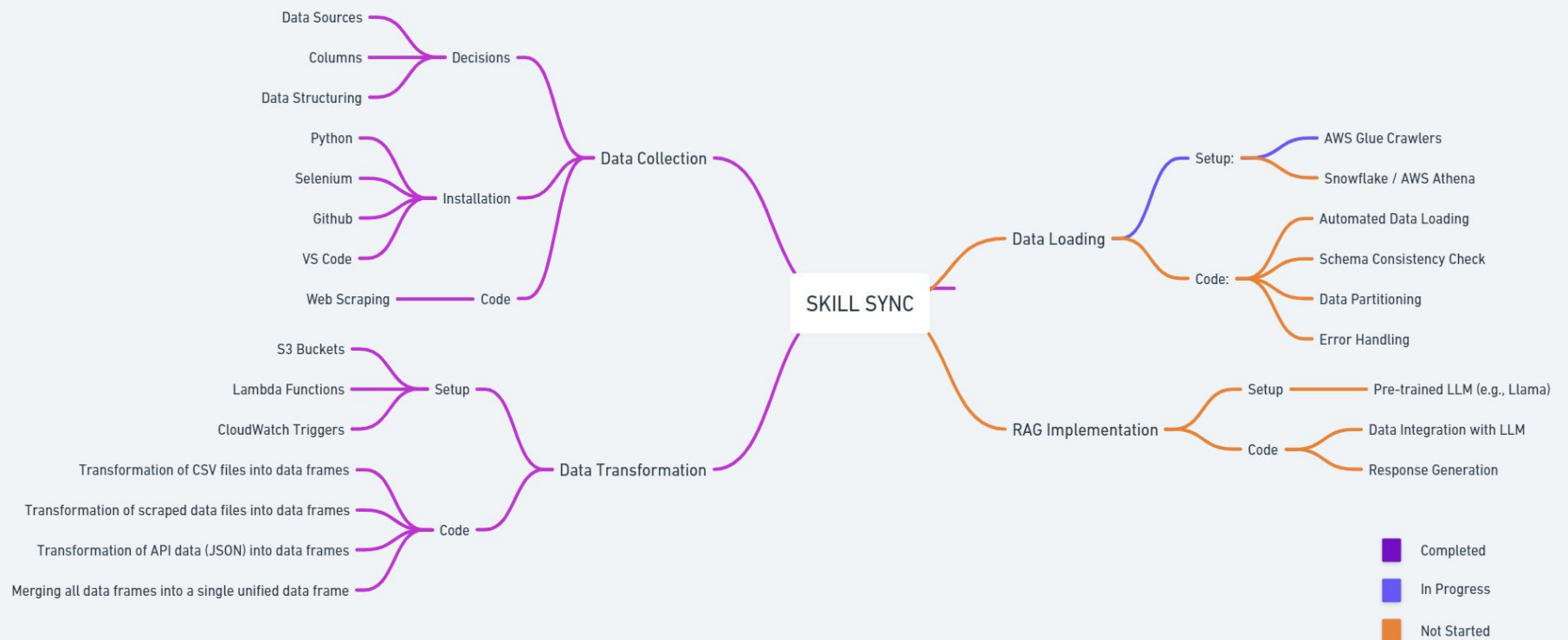Sri Krishna Sai Bhuma

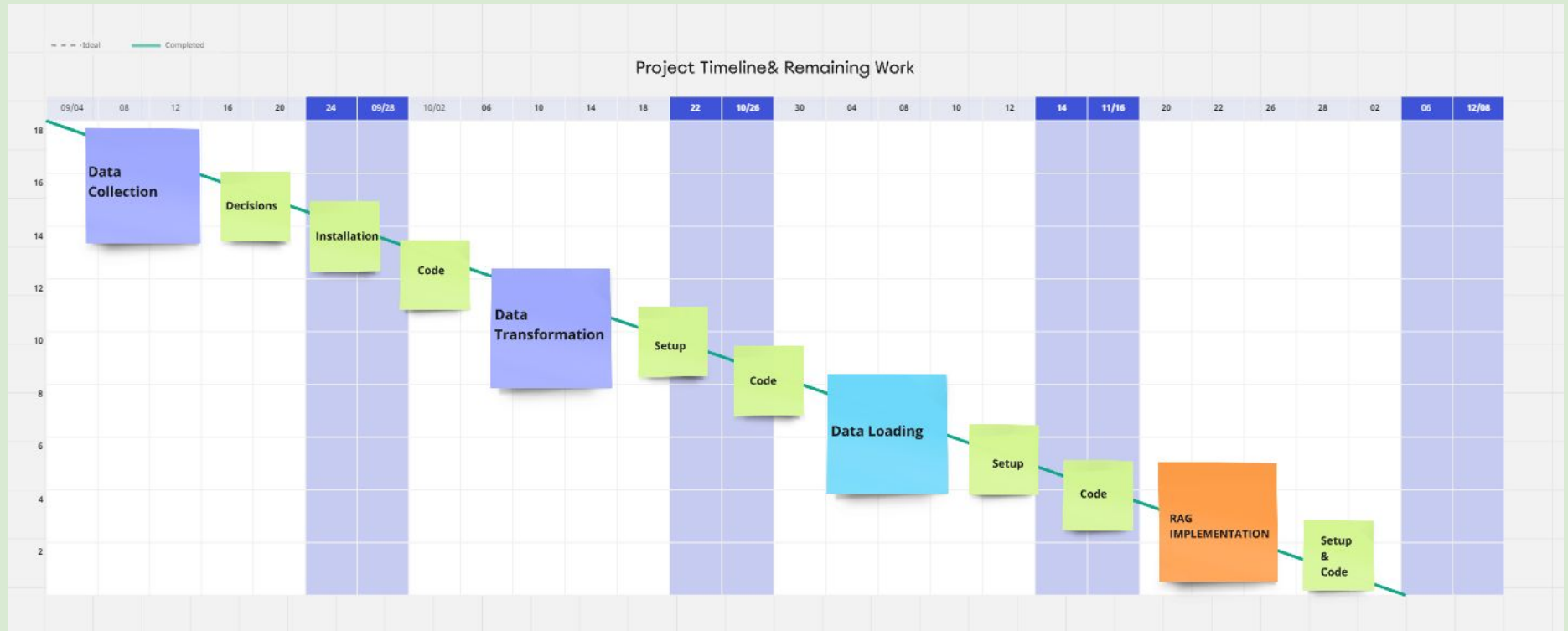# Review of Proposed Deliverables

Core Project Deliverables

- Automated ETL Pipeline on AWS
- Dashboard for User Input
- Final Presentation

Key Technical Components

- AWS (S3, Trigger, Lambda Functions)
- Snowflake
- Python
- Selenium

Project Pipeline

Mind map titled "SKILL SYNC" with branches:

Data Collection
- Decisions: Data Sources, Columns, Data Structuring
- Installation: Python, Selenium, Github, VS Code
- Code: Web Scraping

Data Transformation
- Setup: S3 Buckets, Lambda Functions, CloudWatch Triggers
- Code: Transformation of CSV files into data frames, Transformation of scraped data files into data frames, Transformation of API data (JSON) into data frames, Merging all data frames into a single unified data frame

Data Loading
- Setup: AWS Glue Crawlers, Snowflake / AWS Athena
- Code: Automated Data Loading, Schema Consistency Check, Data Partitioning, Error Handling

RAG Implementation
- Setup: Pre-trained LLM (e.g., Llama)
- Code: Data Integration with LLM, Response Generation

Legend:
- Completed
- In Progress
- Not Started

Current Challenges And Mitigations:
1) Data integration Complexity
2) Technical Instruction Setup

Mitigation Strategy:
1) Weekly Technical Reviews
3) Post-Project RetroSpectives.

THANK YOU